Data Integration Issues in Research Supporting Sustainable Natural Resource Management

ALEXANDER HERR
CSIRO, Davies Lab, PMB PO Aitkenvale, QLD 4814, Australia.
Email: alexander.herr@csiro.au

Received 20 February 2007; Revised 3 May 2007; Accepted 9 May 2007

Abstract
Current decision-making in natural resource use and management aims at delivering ecologically-sustainable development to achieve conservation and economic benefits. The process of guiding natural resource use requires the integration of social, economic and biophysical information on which to base management decisions. This paper discusses the integration of socio-economic information for natural resource management (NRM) planning and decision-making in the Australian context. A comprehensive resource of socio-economic data is the Census, which is undertaken every five years by the Australian Bureau of Statistics (ABS) for the whole of Australia. Unfortunately there are qualitative and quantitative issues stemming from the use of ABS census data maps for NRM decision-making, as they are at a different scale to and the boundaries do not coincide with biophysical information. These issues include the variable shape of collection districts, the use of enumerated data for population-based statistics, the large size of collection districts in low populated areas, and the averaging of socio-economic information over the collection districts. Examples highlight these issues and show a way forwards in improving data integration, which includes simple spatial overlay methods and regression modelling.

KEY WORDS *socio-economic; biophysical; integration; ecologically-sustainable development*

ACRONYMS

ABS	Australian Bureau of Statistics
CD	Collection District
GBR	Great Barrier Reef
NRM	Natural resource management
SLA	Statistical Local Area
URP	Usual residents' profile

Introduction

In the light of ecologically-sustainable development, natural resource management (NRM) in Australia has moved from a mostly biophysical regulatory focus to an approach that integrates socio-economic perspectives and community participation, as recent planning documents, reports and funding initiatives attest (for example, Commonwealth of Australia, 2001; Council of Australian Governments, 1992; Greiner *et al*., 2003; The State of Queensland and Commonwealth of Australia, 2003; Greiner *et al*., 2005). The research required for NRM has also shifted to provide solution-based and outcome-focused

doi: 10.1111/j.1745-5871.2007.00476.x

science. This requires the integration of different research disciplines and a focus on integration of information from the social, economic and biophysical arenas, regionalisation of planning and NRM governance, as well as an emphasis on stakeholder engagement in decision-making and on-ground works (Daly, 1992; Morrison *et al*., 2004; Broderick, 2005; Farrelly, 2005; McAlpine *et al*., 2007).

Advances in interdisciplinary research have provided conceptual understanding and frameworks for sustainable integration of natural resource use with economic systems and society (Daly, 1992; Arrow *et al*., 1995; Ludwig *et al*., 1997). Still, there are significant problems for outcome-focused and targeted regional NRM planning; for example in biodiversity conservation (McAlpine *et al*., 2007), which at a more basic level requires the spatial and temporal integration of information for NRM and policy development, and finding a common denominator applicable on-ground. It is the aim of this paper to identify spatial data integration issues in the Australian context and to provide examples overcoming the data mismatch in the NRM arena.

The spatial dimension of integration faces several difficulties in terms of scale mismatch, lack of overlap for specific locations, and a mismatch in the temporal scale of the data sets. For example, economic reporting is generally non-spatial (such as per capita) or covers an area unrelated to natural resource extraction, while ecological processes and relationships show high spatial dependence and variations in temporal scale. A further complication is the difference between the (larger) ecological and economic timescales (Levin, 1992; Jordan and Fortin, 2002; Trewin, 2003).

However, there are options to integrate biophysical and socio-economic data by overlaying (to achieve 'best match'), averaging, aggregating, disaggregating and modeling. This paper discusses the first and the last options using examples from Outback (in this context meaning sparsely populated) Australia. Its aims are to:

1. identify data issues relevant to current NRM planning and NRM supporting research in the Australian context, particularly over large Outback areas; and
2. provide examples and specific case study solutions. These solutions are intended to provide examples for planners, policy makers and scientists faced with providing outcomes for ecologically-sustainable development of natural resources.

Data issues: population statistics for the rangelands and an Outback NRM area

While there is a range of biophysical data available at different scales from State, Territory and Commonwealth agencies, obtaining spatially-referenced, socio-economic data is more difficult. The most readily-available data source is Census information produced every five years by the Australian Bureau of Statistics (ABS) (http://www.abs.gov.au). This paper only discusses examples from the ABS data, mainly the CDATA and Integrated Regional Database (2001 Census data products available from the ABS), because they provide Australia-wide (that is, State and Territory overarching) coverage. This section describes the spatial and numerical data limitations associated with the integration of socio-economic and biophysical data.

The finest resolution of Census data comes at the collection district (CD) level. Several CDs combined together form a statistical local area (SLA) and these in turn are the basis for a local government area. Spatial limitations in Outback areas are a result of the sparse settlement and the census collection procedures. There are several issues associated with this for Outback areas. The example in Figure 1 shows that the Lake Eyre Basin boundary for most of its way intersects and crosses collection district boundaries, and also shows the large size of collection districts in remote Outback regions.

The census generally takes place during winter holidays (August/September) which results in some people not being at their normal place of residence. Figure 2 shows the numbers of visitors at the 2001 census and highlights the high proportion in Outback areas, with several CDs having more than 20% and some even more than 75% visitors. Statistics based on enumerated data become an issue when there is a high proportion of visitors and the socio-economic information presented includes these.

Table 1 summarises the numerical data issues associated with this way of aggregating household information. Figure 1 serves as an example, where the crosshatching indicates the areas for which the number of residents counted at home during the census was below 75%.

A high visitor proportion also implies that statistics stemming from enumerated populations will provide a distorted picture of the actual situation. For example, some NRM issues and planning require an adaptive approach when dealing with indigenous stakeholders (Larson *et al*., 2006). However, there are numerical issues when

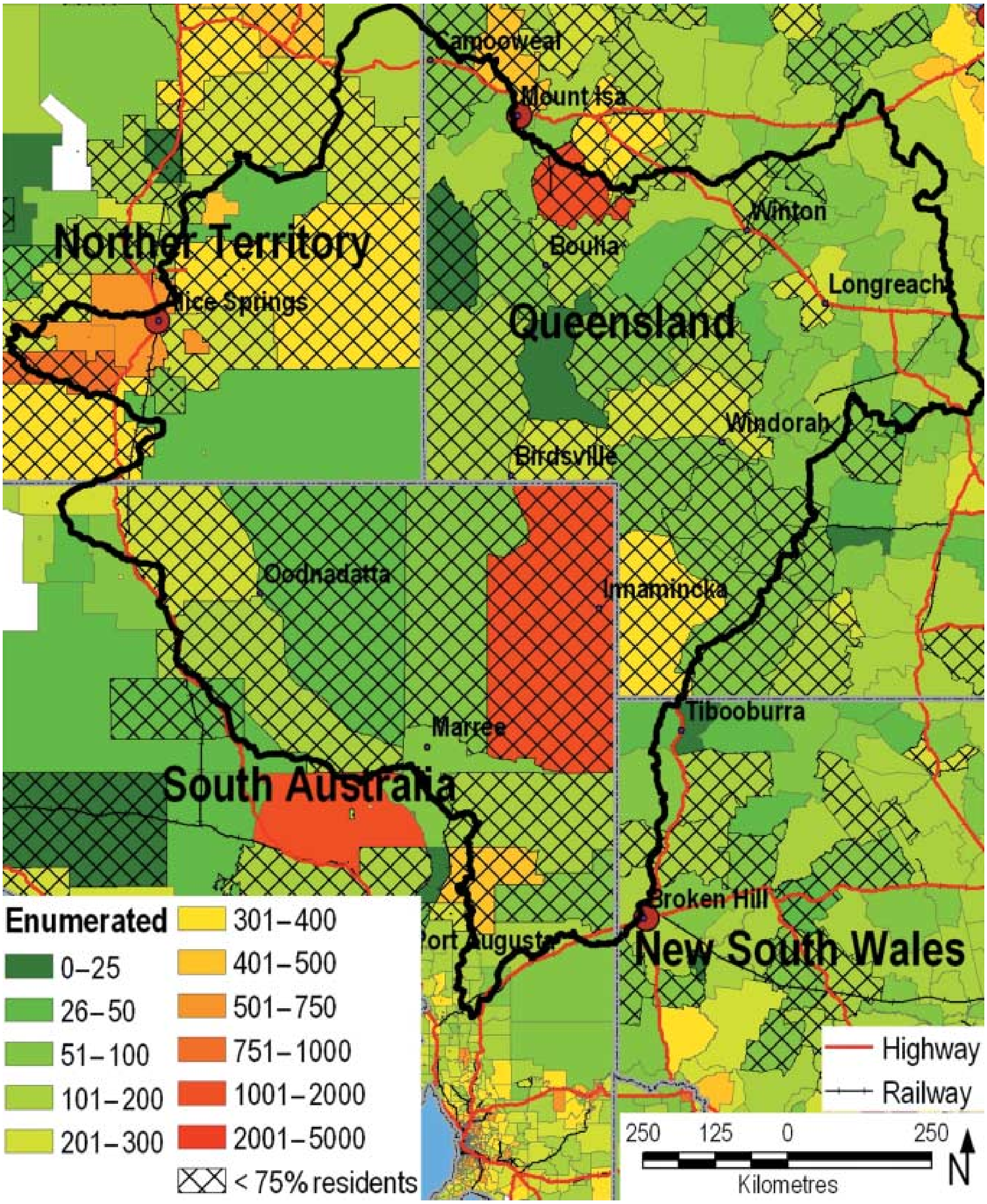

Figure 1 Enumerated population and residents counted at home at CD level for the Lake Eyre Basin NRM area (black line).

trying to establish the proportion of indigenous persons related to enumerated data. Up until the 2006 Census, the ABS presented population information as enumerated data available at the CD level. Usual resident data, which are more suitable for this task, have the visitor component removed and come at the SLA level from the usual residents' profiles (URPs). In the Outback, the SLAs are large and so the URP data are of limited use for research at the regional level. Many small communities occur individually as their own CD (lower part of Figure 3). The data processing approach of the ABS aggregates and averages over the whole area of an SLA for the

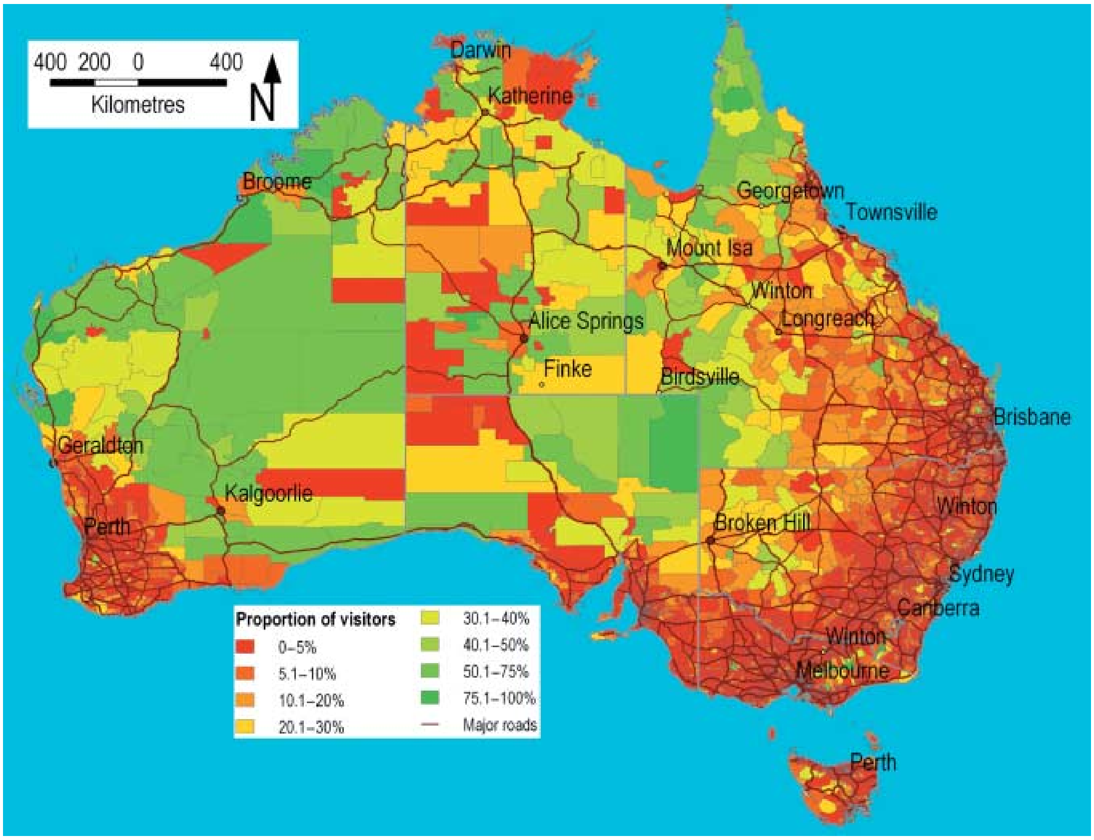

Figure 2 Visitor proportion (CD level) in Australia for the 2001 census.

Table 1 Spatial and numerical data issues of ABS census data at collection district level (finest resolution).

Data issue	Issue	Consequence
Spatial	CD has variable shape between Census years	CD information not directly comparable between Census years
	Information averaged over whole CD area	Aggregation over large areas masks local information and biophysical information of comparable scale
	CD does not cross any other boundaries of the ASGC (Edwards, 2001) and aggregates cover the whole of Australia (Herr and Stoeckl, 2003)	Biophysical information of comparable scale rarely available
	Outback CD often covers several 100 km2	Large area CD boundaries do not match readily with biophysical boundaries
Numerical	Prior to 2006 Census, CD data came in enumerated form	Non-residents are included in the household-specific information, thus masking the real population with visitors and non-inclusion of residents away from home at Census night
	Where the number of variables in the census tables could allow identification of individuals, ABS introduces randomisation into the data to prevent this	Randomisation introduces additional error for population estimates

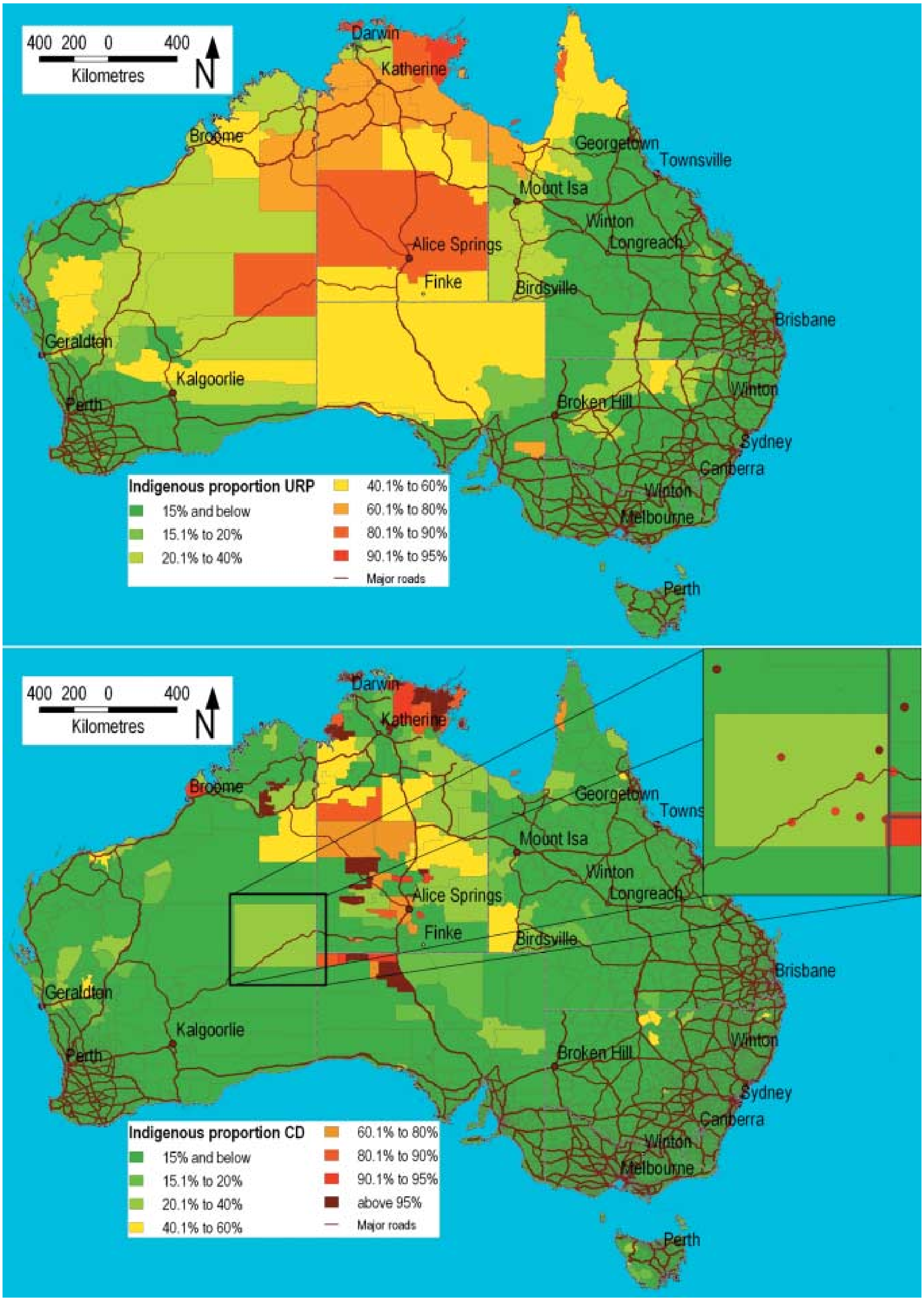

Figure 3 Comparison of CD-level data with aggregated usual residents' profile (URPs) at SLA-level data.

URP. This leads to markedly different proportions when visually comparing indigenous proportions from URPs (upper part of Figure 3) with the (enumerated) indigenous proportion at CD level.

Reasons for these differences in the two images of Figure 3 are:

1. randomisation of CD level data for confidentiality purposes;
2. averaging over a large area, thus losing the finer resolution (URPs at SLA level);
3. inclusion of visitors into the basic community profiles information (CD level); and
4. in this specific context, the size of categories may marginally influence the difference (that is, if differences between SLAs and CDs lie within one category).

Additionally, for areas with large indigenous populations, there are issues related specifically to the collection of census data in indigenous communities, which may lead to misinterpretation of family membership, resulting in collectors under- or over-counting residents. This can reduce the reliability of the information with results differing between the various statistical levels (Martin and Taylor, 1995; Martin *et al*., 2002).

Examples of data integration

In the subsequent part of this section Australia's rangelands and the Lake Eyre Basin NRM area provide examples of the challenges encountered when attempting to provide population statistics for NRM areas in the Outback. Data issues in this context are:

1. mismatch of biophysical and socio-economic boundaries;
2. quality, and
3. resolution of the available socio-economic data.

The previous section outlined the data format of socio-economic census data. While it is crucial for modern NRM to integrate biophysical and socio-economic information, the socio-economic data format (and specifically the boundaries) do not coincide with areas in which NRM takes place. Solutions for integrations include; a simple overlay with visual assessment ('best match'), and modeling to match economic values to NRM boundaries.

'Best match' overlay of biophysical and socio-economic data

The following example is from the Burdekin Dry Tropics natural resource management process. The Burdekin Dry Tropics Board, the NRM Body charged with developing the NRM plan for the Burdekin Dry Tropics under the National Action Plan for Water Quality and Salinity, required socio-economic information for the Burdekin catchment. This meant matching ABS data to these NRM boundaries. The data are at a sufficient scale for which simple overlay and visual estimation are satisfactory at the CD level (Greiner *et al*., 2003). A basic GIS overlay procedure of the different data set can achieve this. Figure 4 and 5 are examples of such an overlay. The index of socio-economic disadvantage shown in Figure 4 provides a normative measure of socio-economic status within an area with high scores, indicating few families with low income, low education, little training and unskilled occupations (Trewin, 2001). Here the overlay was sufficient to obtain an overview of the socio-economic status for the Burdekin NRM area,

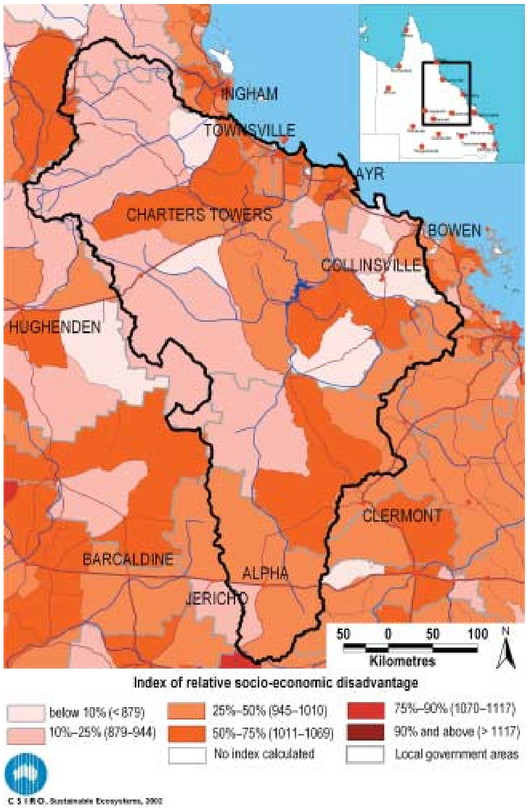

Figure 4 Index of socio-economic disadvantage in the Burdekin Dry Tropics NRM area for 1996. A high index means a low disadvantage (Trewin, 2001). CDs overlapping the Burdekin Dry Tropics NRM area (outlined by the heavy line), obtained by overlaying the ABS data on the NRM area (Greiner *et al*., 2003).

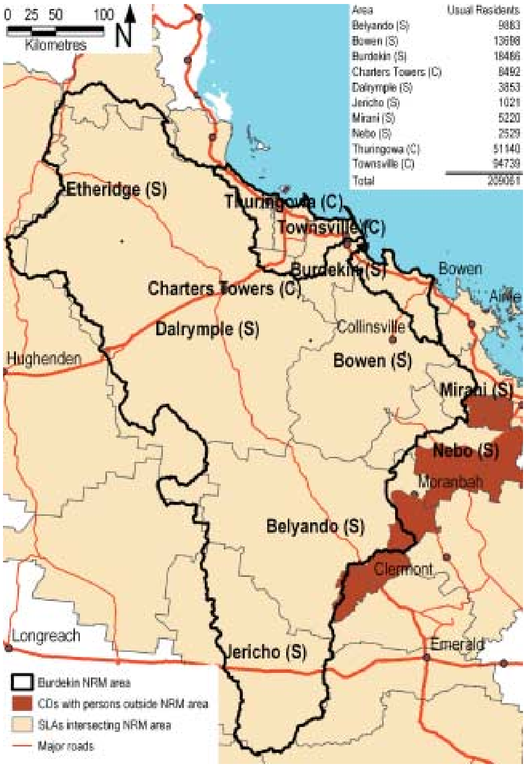

Figure 5 Population in the Burdekin Dry Tropics NRM area (Census year 2001) based on the SLAs, obtained by overlaying the socioeconomic data on the NRM area but discounting the persons counted in the CDs (dark areas in image) outside the area of interest.

which in 1996 was lower for most of the areas than the national average (Greiner *et al*., 2003).

However, when attempting to estimate the population in the Burdekin NRM area, some improvements are required. For example, a visual assessment identifies the SLAs best matching the area for estimating the usual residents (area listing in Figure 5 top right). Further improvements are possible when subtracting the persons counted at home (in lieu of usual residents data) from the CDs (as indicated in Figure 5) within these SLAs but outside the NRM area. This brings the number of persons living in the NRM area to approximately 200 000 for the Census year 2001 (see also Greiner *et al*., 2003).

Modeling of economic values on a catchment basis

Regional NRM and policy development requires socio-economic information at the catchment level. Agricultural-economic information from the ABS comes at the SLA level and is not of sufficient resolution for simple overlay with catchment information in northern Australia. Alternatives include disaggregating (and using averages of) economic data or modeling. The following provides an example of modeling for integration with a discussion of associated issues and options for improvement.

The Great Barrier Reef (GBR) World Heritage Area receives runoff from a large number of catchments (37) in Northern Australia. Agricultural activity increases pollutants and sediment transport through this runoff to the GBR. Recent government initiatives have identified this as a major issue, and solutions need to address it in a biophysical and social context. The development of the Reef Water Quality Protection Plan incorporated biophysical and socio-economic dimensions for use in management prioritisation (Greiner *et al*., 2005). This process included the generation of agricultural production values for each GBR catchment data at the SLA level (ABS, 2002). Modeling as a first attempt in producing these values was a three-step process that linked land-use information (Steward *et al*., 2001) to a three-tiered level of agricultural production. The images in Figure 6 show this process in a stylised form.

These three steps are:

1. establish land-use categories with low, medium and high value agricultural production areas;
2. a) in the SLA identify the area of each land-use category;
   b) statistically model agricultural production values for each land use based on regressing the area of each categorised agricultural production area against the total agricultural production value for each SLA (Figure 7); and
3. sum the agricultural production values per land-use category in each catchment.

Figure 8 shows the outcomes of this modeling.

There are discrepancies in the modeling when comparing the predicted with the SLA values in the coastal areas, where sugar and other high value crops occur. These differences stem from five different error sources.

1. Land-use data resolution is broad (approximately 1 km^2), so some (high value) crops grown in smaller areas (namely beyond the accuracy of the data collection) are not included.
2. Temporal discrepancies between the land-use data from 1996/1997 and the Agricultural

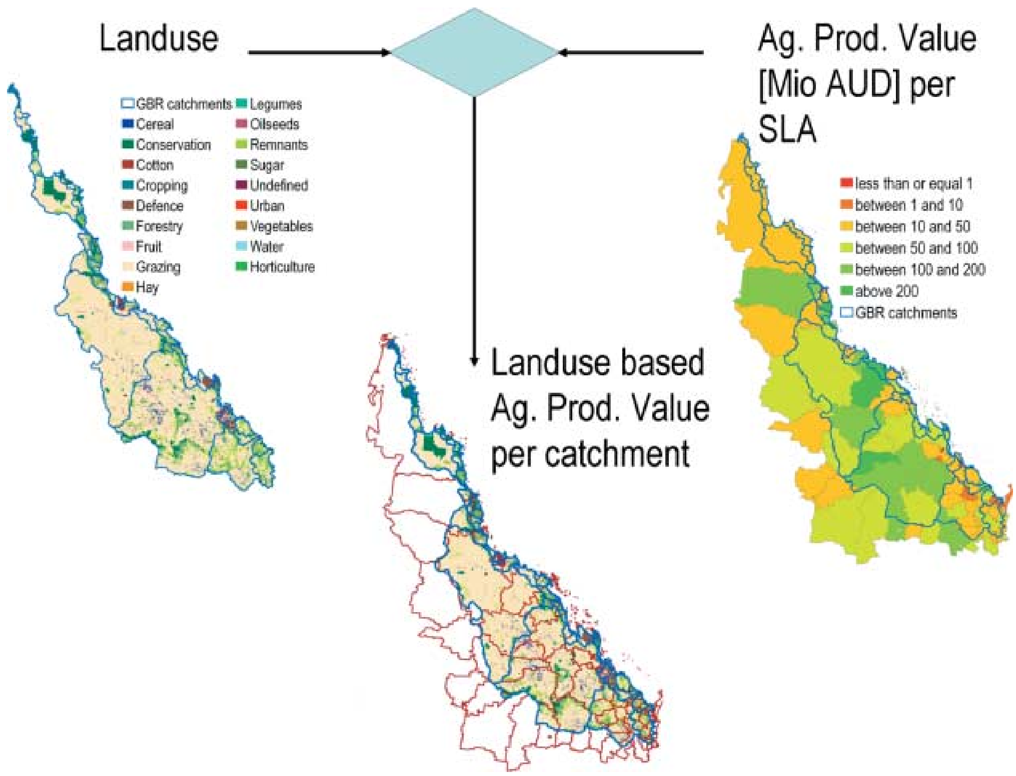

Figure 6 Schematic overview showing the modeling of Agricultural production value on a catchment basis.

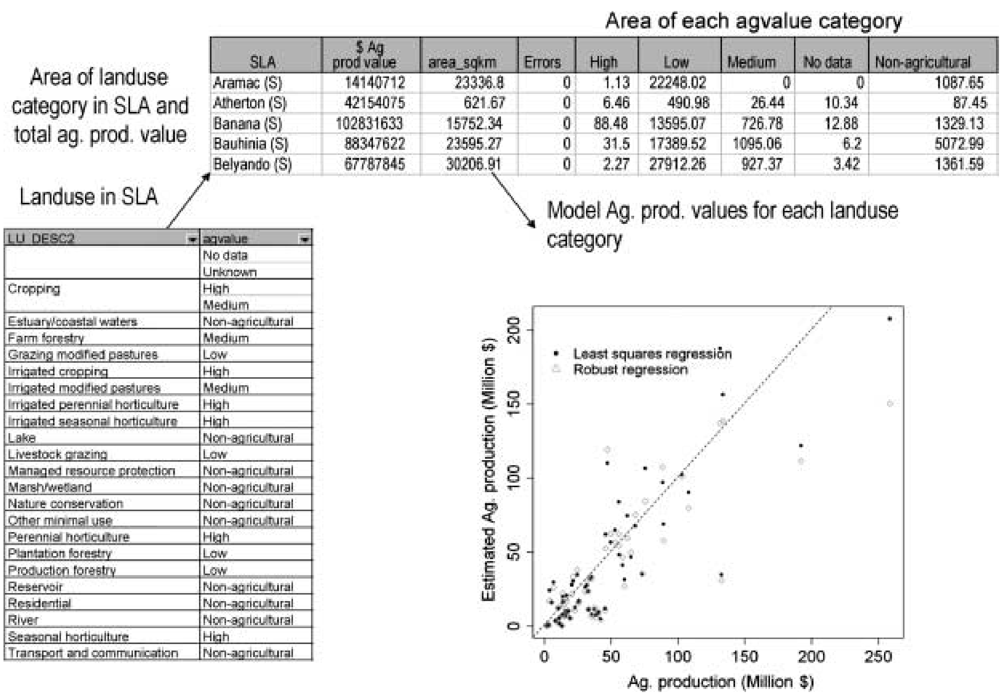

Figure 7 Modeling process for estimating agricultural production value on a catchment basis (see text for explanation).

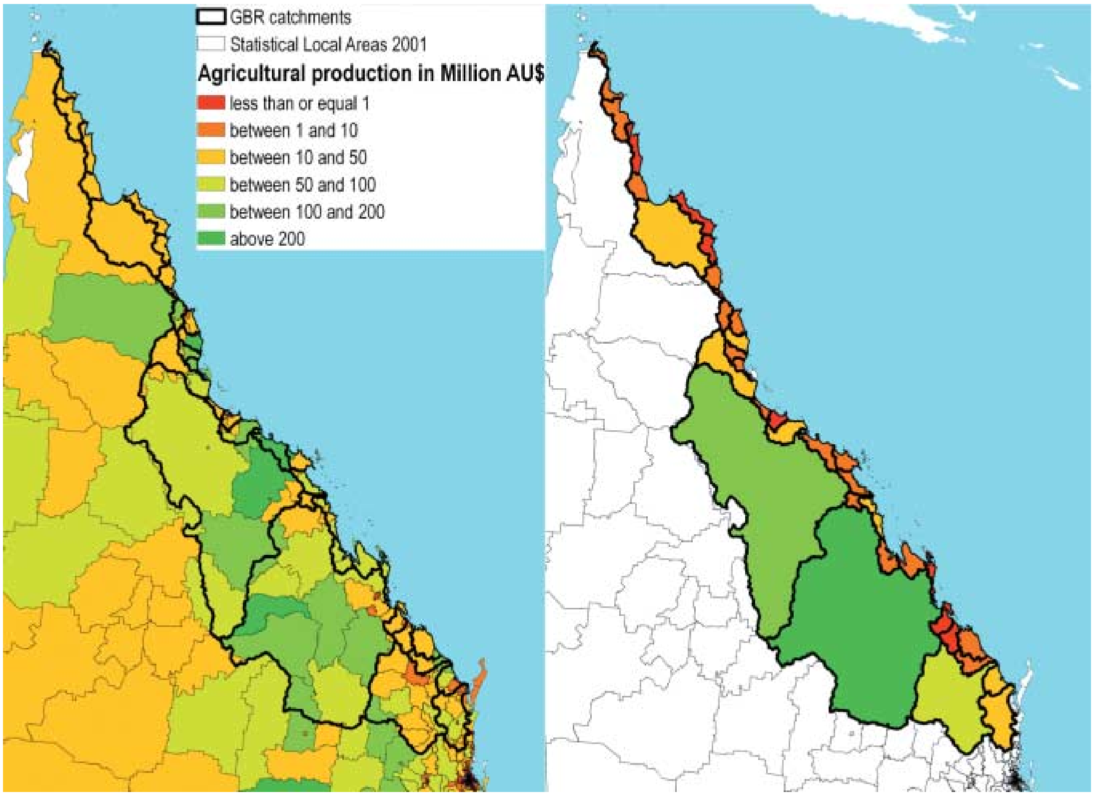

Figure 8 Modeled agricultural production value in the GBR catchments (right) and agricultural values per SLA (left).

Production values from 2001 (ABS, 2002) may result in errors through land-use change within the time frame.

3. There are errors in the land-use classification and the model does not include no-data values (that is, uncertainty about the real land-use class).
4. The model assumes a linear relationship of land use with agricultural production.
5. There may be a (spatial) correlation between the low, medium and high land-use categories.

Improving the modeling would produce a more accurate prediction of agricultural production values per catchment. These would include data quality improvements (that is, more accurate land-use data from the same term as the Agricultural Production Values) and improvements in the modeling method. Modeling improvements could include outlier detection, identification and accounting for non-linear estimation, closer alignment of land use with crop-specific agricultural values, and providing information/mapping on uncertainty stemming from the modeling (for example, a map of range or standard error for predicted data).

The previous section discussed two approaches for integrating socio-economic and biophysical data over large areas of Australia, where the knowledge and information resources are limited and there is a mismatch of time, scale and boundaries of the different data sources (Herr and Stoeckl, 2003; Stoeckl and Stanley, 2005). As the first example shows, data averaging and visual assessment to achieve a 'best match' is the simplest means of data integration. Although there are more advanced spatial methods available (for example Luo, 2004), this is likely to be sufficient when determining population statistics for catchment-based NRM planning, particularly in areas of low population density such as the Outback.

While a 'best match' approach is intuitive and sufficient for simple overlays of biophysical and socio-economic data, it becomes more unreliable where the visual assessment requires complementation through intensive user input. From here on, transparency and objectivity may become a critical issue as it is difficult to reproduce a person's individual judgement. In these instances a modeling approach may be more acceptable if there is documentation and justification of the modeling type.

In the United States, researchers, who combined socio-economic data from the Census with remote-sensing information to provide integration at the catchment level (Kuczenski *et al*., 2000; Radeloff *et al*., 2000), identified that the relationships between these data are complex and concluded that there is a need to increase knowledge about the functioning of rural societies in relation to their resource use. Others argue that census data may well be insufficient for watershed managers, and primary, targeted data collection through surveys is necessary for successful management at catchment level (Curtis *et al*., 2005). However, modeling of biophysical and census data can yield an initial overview that is sufficient to direct further research and data collation, and where the purpose is to provide an initial snapshot of the socio-economic status at catchment level.

Conclusion

Most of Australia has very low population densities and sparse data, and socio-economic and biophysical boundaries do not match. This provides challenges for the integrated (that is, socio-economic and biophysical) assessment and management of NRM issues. Issues for integration are the quality of the data (for example, data collection in remote communities) and the method of data presentation (for example, enumerated data *versus* usual residents). This becomes particularly important in Outback areas where visitation during the Census period (winter and holiday season) is high. Here, NRM planners using these data should at least be aware of the number of visitors in the data they use for their analyses and factor these into their interpretations even if they make use of 'best match' or modeling options.

ACKNOWLEDGMENTS

Funding from the Burdekin Dry Tropics Board, the Desert Knowledge CRC and the CSIRO Water for a Healthy Country Flagship supported the examples illustrated in this paper.

REFERENCES

ABS, 2002: *Integrated Regional Data Base*. Catalogue No. 1353.0. Australian Bureau of Statistics, Canberra.

Arrow, K., Bolin, B., Constanza, R., Dasgupta, P., Folke, C., Holling, C.S., Janssen, B.-O., Levin, S., Mäler, K., Perrings, C. and Pimentel, D., 1995: Economic growth, carrying capacity, and the environment. *Science* 268, 520–521.

Broderick, K., 2005: Communities in the catchments: implications for natural resources management. *Geographical Research* 43, 286–296.

Commonwealth of Australia, 2001: *Australia's Salinity Problem*. The Australian Government, Canberra.

Council of Australian Governments, 1992: *National Strategy for Ecologically Sustainable Development*. Australian Government Publishing Service, Canberra.

Curtis, A., Byron, I. and Mackay J., 2005: Integrating socio-economic and biophysical data to underpin collaborative watershed management. *Journal of the American Water Resource Association* 41, 549–563.

Daly, H., 1992: Allocation, distribution and scale: towards an economics that is efficient, just and sustainable. *Ecological Economics* 6, 185–193.

Edwards, R.W., 2001: *Australian Standard Geographical Classification*. Australian Bureau of Statistics, Canberra.

Farrelly, M., 2005: Regionalisation of environmental management: a case study of the Natural Heritage Trust, South Australia. *Geographical Research* 43, 393–405.

Greiner, R., Herr, A., Brodie, J. and Haynes, D., 2005: A multi-criteria approach to Great Barrier Reef catchment (Queensland, Australia) diffuse-source pollution problem. *Marine Pollution Bulletin* 51, 128–137.

Greiner, R., Stoeckl, N., Stokes, C., Herr, A. and Bachmaier, J., 2003: *Natural Resource Management in the Burdekin Dry Tropics: Social and Economic Issues*. CSIRO Sustainable Ecosystems, Resources Futures, Townsville, Queensland.

Herr, A. and Stoeckl, N., 2003: *A Regional Profile of the Tropical Savannas*. CSIRO, Alice Springs.

Jordan, J.G. and Fortin, M.J., 2002: Scale and topology in the ecological economics sustainability paradigm. *Ecological Economics* 41, 361–366.

Kuczenski, T.K., Field, D.R., Voss, P.R., Radeloff, V.C. and Hagen, A.E., 2000: Integrating demographic and Landsat (TM) data at a watershed scale. *Journal of the American Water Resources Association* 36, 219–228.

Larson, S., Herr, A. and Greiner, R., 2006: Human well-being and natural environments: an indigenous perspective. *The International Journal of Environmental, Cultural, Economic and Social Sustainability* 2, 39–50.

Levin, S., 1992: The problem of pattern and scale in ecology. *Ecology* 73, 1943–1967.

Ludwig, D., Walker, B. and Holling, C.S., 1997: Sustainability, stability and resilience. *Conservation Ecology* [online]1(1): 7. Available from the Internet. URL: http://www.consecol.org/vol1/iss1/art7/

Luo, W., 2004: Using a GIS-based floating catchment method to assess areas with shortage of physicians. *Health and Place* 10, 1–11.

Martin, D.F., Morphy, F., Sanders, W.G. and Taylor, J., 2002: *Making Sense of the Census: Observations of the 2001 Enumeration in Remote Aboriginal Australia*. 22. Centre for Aboriginal Economic Policy, Australian National University, Canberra.

Martin, D.F. and Taylor, J., 1995: *Enumerating the Aboriginal Population of Remote Australia: Methodological and Conceptual Issues*. CAEPR Discussion Paper No. 91. Centre for Aboriginal Economic Policy, Australian National University, Canberra.

McAlpine, C.A., Heyenga, S., Taylor, B., Peterson, A. and McDonald, G., 2007: Regional planning in Queensland's Rangelands: challenges and prospects for biodiversity conservation. *Geographical Research* 45, 27–42.

Morrison, T.H., McDonald, G.T. and Lane, M.B., 2004: Integrating natural resource management for better environmental outcomes. *Australian Geographer* 35, 243–258.

Radeloff, V.C., Hagen A.E., Voss, R.R., Field, D.R. and Mladenoff, D.J., 2000: Exploring the spatial relationship

between census and land-cover data. *Society and Natural Resources* 13, 599–609.
Steward, J.B., Smart, R.V., Barry, S.C. and Veitch, S.M., 2001: *1996/97 Land Use of Australia – Final*. National Land and Water Resources Audit, Canberra.
Stoeckl, N. and Stanley, O., 2005: Land rich and data poor: modelling requirements in Australia's far north. *Economic Papers* 24, 230–248.
The State of Queensland and Commonwealth of Australia, 2003: *Reef Water Quality Protection Plan; for Catchments Adjacent to the Great Barrier Reef World Heritage Area*. Department of the Premier and Cabinet, Brisbane.
Trewin, D., 2001: *Socio-economic Indexes for Areas*. Australian Bureau of Statistics, Canberra.
Trewin, D., 2003: *Australian Standard Geographical Classification (ASGC) 2003*. Australian Bureau of Statistics, Canberra.